

Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | SWTID1720078183 |
| Project Title | Predictive Modeling for Fleet Fuel Management using Machine Learning |
| Maximum Marks | 2 Marks |

**Data Quality Report**

This document outlines the data quality issues encountered in the fuel consumption dataset and provides a plan for addressing each issue to ensure the accuracy and reliability of the data used for predictive modeling.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
|  | Missing values in fuel consumption data | Moderate | Impute missing values using median values  Remove duplicates based on unique transaction ID  Identify and remove outliers using IQR method  Identify and remove outliers using IQR method |
| Fuel | Duplicate entries | High |
| Consumpti |  |  |
| on Records | Outliers in fuel consumption amounts | High |
|  | Outliers in speed data | High |